

## **REVIEW: “Competitive intelligence and patent analysis in drug discovery: Mining the competitive knowledge bases and patents”**

Grandjean, N., Charpiot, B., Pena, C.A., Peitsch. M.C (2005)

(Recensión por M. Welling Flensburg)  
e-mail: [m.welling.flensburg@gmail.com](mailto:m.welling.flensburg@gmail.com)

El artículo pretende establecer la relación que existe entre el análisis de patentes y la inteligencia competitiva en el contexto de la industria farmacéutica, en particular en el proceso de desarrollo de fármacos.

El texto analiza cómo las técnicas de análisis de patentes y de exploración de bases de datos se aplican y como estos métodos pueden ayudar en la tarea de conseguir información sobre las actividades de la competencia, actividades de I+D que se están realizando, campos con potencial de crecimiento, o sobre las colaboraciones que se están haciendo en este campo. Los autores indagan en las ventajas y las desventajas del análisis de patentes, aunque concluyen que las patentes siguen siendo la fuente más fiable de información sobre actividades de I+D.

En primer lugar, los autores evidencian el retraso temporal que existe entre el momento de la inscripción de una innovación a las oficinas de patentes, y el momento en que se publican estas patentes. Este retraso, de unos 18 meses, representa un desfase entre el estado de la técnica "real" y lo que se publica, lo cual representa un inconveniente. A pesar de este problema, el análisis de patentes o Patentometría es el mejor recurso existente para conocer el estado de las actividades de I+D de una industria determinada. Dada la importancia de los registros de patentes, su calidad determina en gran medida la utilidad que se le pueda dar (especialmente en los casos en los que se quiera conocer mejor la estrategia de una empresa basándose en las patentes que haya registrado). En este contexto, el principal problema de los archivos de patentes primarios (aquellos que recogen lo que es publicado por las oficinas de patentes) radica en su complejidad general que dificulta una interpretación estratégica. En lo que respecta a las bases de datos comerciales, indexadas por terceros, su principal problema es la falta de detalle y profundidad en los contenidos.

A continuación, los autores explican cómo, ante las dificultades para interpretar sistemáticamente las bases de datos existentes, puede ser interesante realizar un proceso de extracción de información útil de las patentes. El proceso en cuestión puede ser relativamente sencillo, en los casos en los que haga falta detectar y normalizar conceptos de forma más o menos sistemática, o puede ser más complejo. En este segundo caso, la extracción requiere métodos más complejos de procesamiento de texto, que permitan obtener un conjunto de datos susceptibles de análisis, exploración e interpretación posterior. Un buen ejemplo de extracción compleja sería el caso de conceptos como 'nombre de la empresa', 'autores de la invención' o 'modelos de actuación' que pueden ser extraídos y utilizados por más de una base de datos de segundo nivel, en función del criterio que rijan cada base de datos en cuestión. De esta manera, estos conceptos pueden utilizarse a modo de conectores entre estas bases de datos secundarias dentro del archivo principal.

Los principales pasos del proceso de extracción de información de las patentes son:

Nº de fase	Fase
1	<b>Zoning:</b> <i>marcar frases relevantes identificadas según su importancia científica.</i>
2	<b>Parsing:</b> <i>extracción de las frases en un archivo de datos</i>
3	<b>Normalización:</b> <i>se hace aplicando reglas (algoritmos) o bien asignando términos a los sinónimos</i>
4	<b>Extracción de conceptos:</b> <i>desde el texto, teniendo en cuenta el archivo de datos creado en los pasos 2 y 3, es decir, que extrae conceptos considerando los sinónimos que pueden adoptar los conceptos y el contexto en lo que se encuentran</i>
5	<b>Categorización:</b> <i>ordenación de los conceptos en sub-archivos</i>
6	<i>Aplicación de un algoritmo de taxonomía que permita navegar por la base de datos en coherencia con el criterio de búsqueda que se desee</i>
7	<b>Relacionar conceptos:</b> <i>las relaciones se establecen aplicando reglas de análisis sintáctico.</i>

El proceso anterior permite crear bases de datos con funcionalidad y sensibilidad a los criterios de investigación que sean de interés para el usuario. Estas bases de datos tienen que estar actualizadas y tendrán que ser científicamente coherentes, matemáticamente válidas, y susceptibles de poder ser exploradas de forma funcional.

Para explorar y analizar las bases de datos, se suelen utilizar técnicas específicas de exploración, que permitan establecer relaciones entre distintas variables concretas. Además se suelen utilizar también herramientas gráficas que ilustran mejor las relaciones entre variables y también son capaces de detectar nuevos patrones inéditos u otras informaciones que antes no se conocían.

Por último, el artículo se cierra con las alternativas comerciales disponibles, tanto productos para el desarrollo de análisis de patentes y exploradores de bases de datos, como paquetes comerciales ya desarrollados.